Received: 24 May 2022 | Revised: 10 October 2022 | Accepted: 21 October 2022

DOI: 10.1002/pds.5564

ORIGINAL ARTICLE

WILEY

Social media for early characterization of pandemic symptoms: A qualitative analysis of patient-reported COVID-19 experiences

Melissa Khashei[1] | Scott Janiczak[2] | Christopher St. Clair[3] | Wei Liu[2] | Jae Joon Song[4] | Wei Hua[2] | Monique Falconer[2] | Efe Eworuke[2]

[1]Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, USA

[2]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, USA

[3]Office of New Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, USA

[4]Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA

Correspondence
Efe Eworuke, Division of Epidemiology II, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA.
Email: efe.eworuke@fda.hhs.gov

Abstract

Background: Patients use social media forums to discuss their medical history and healthcare experiences, providing early insight into real-world patient experiences. We analyzed COVID-19 patient experiences from Reddit social media posts.

Methods: We extracted Reddit Application Programming Interface data for the subreddit/COVID-19 positive from March to August 2020 and selected users tagged as "Tested Positive" or "Tested Positive- Me" flair and who posted at least thirty times in any calendar month, excluding users who explicitly stated location outside of the U.S. For tested-positive patients (users), we created and reviewed individual case profiles summarizing their COVID-19 symptoms, testing, and medications or treatments. Data were imported to Nvivo qualitative analysis software and qualitative coding was conducted.

Finding: There were 31 759 posts and comments from 720 users in March to May 2020 (Q1) and 40 446 posts and comments from 1649 users from June to August 2020 (Q2). Final count of "Tested Positive" was 1296 users (280 in Q1 and 1016 in Q2). Across both quarters, frequently reported symptoms included sore throat, headaches, fevers, or chills. Loss of sense of smell or taste were reported by users in early March, prior to the inclusion of this symptom to the CDC list in April and GI-related symptoms and fatigue were reported in the March to May data, before they were added as a COVID-19 associated symptom in July 2020. Users also reported in-depth descriptions of their symptoms, motivations for testing, and long-term impacts such as post-viral fatigue.

Interpretation: Social media data can potentially serve as an early surveillance data source in a pandemic and offer preliminary insights into patient disease experiences.

KEYWORDS
COVID-19, patient engagement, qualitative, Reddit, research, social media

Key Points

- Reddit is a social media platform that allows users to anonymously discuss their medical history and healthcare experience.

Disclaimer: This presentation represents the views of the user and does not represent FDA official position or policy.

- We analyzed content about COVID-19 symptoms on Reddit in the first two quarters of the pandemic.
- Users reported loss of sense of smell or taste, GI-related symptoms and fatigue in early March, prior to the inclusion of these symptoms to the Centers for Disease Control (CDC) list in April and July 2020.
- Social media data can provide early insights into patient disease experience.
- In-depth discussions of symptoms on social media platforms can also aid in broader understanding of the natural history of the disease and a clinical presentation of symptoms from the patient's perspective.

Plain Language Summary

In this study, we extracted Reddit social media posts for users tagged as COVID-19 positive to explicitly describe their COVID-19 symptoms, testing, medications, or treatments. We extracted Reddit API data from 1296 users who posted comments on Reddit between March and August of 2020. In-depth review of case profiles indicated that users were reporting classic symptoms of COVID-19 such as throat pain, headaches, fevers, chills, loss of taste or smell prior to the naming of these symptoms by the CDC. Users also reported in-depth descriptions of their symptoms, motivations for testing, and long-term impacts such as post-viral fatigue. Social media data can potentially serve as early disease surveillance data source capable of providing preliminary insights into patient disease experiences.

1 | INTRODUCTION

Coronavirus Disease 2019 (COVID-19) pandemic, for the first time created an unprecedented wide-spread and active interaction on social media platforms.[1] As with emerging diseases, patients will often research symptoms and alternative therapies before seeking medical advice. In the case of COVID-19, online support forums became avenues for patients to share their experiences about the disease symptoms and treatment options, and even medical advice from other patients. A review of publications on COVID-19 and social media during the first outbreak identified several themes.[2] Social media was used to survey public attitudes, identify infodemics, assessing mental health, detecting, or predicting COVID-19 cases, analyzing government responses to the pandemic and evaluating the quality of health information in prevention education videos.[2] Beyond these, social media can also be used to learn about patients' disease experience.[3] By analyzing users' conversations on social media, researchers can gain understanding of the disease including coping strategies,[4,5] identify concerns about the disease and what barriers exist to access to and quality of care.[6] In the case of COVID-19, social media could have presented a unique opportunity to understand the clinical course of the disease early on in the pandemic. Published reports early on in the pandemic were largely drawn from high risk populations such as hospitalized patients,[7,8] pregnant women[9] or from a convenience sample[10] limiting the generalizability of the study's results. These studies also faced the challenges of prospectively recruiting patients which can be expensive and time-consuming. For patients with mild disease who were often asked to self-quarantine at home, social media may enhance the understanding of patient experience beyond what can be learned from traditional sources of data.

Different social media platforms and communities have different standards for how users are identified, how they can interact, what type of information and content can be shared, and how that content is presented. Reddit users can be completely anonymous with only some or no self-reported identifying information allowing people to engage and share information with privacy. Any user can create a community, or subreddit, and moderate content. Subreddits can range in topic, size, and engagement, and many subreddits can exist that focus on the same general concept. Reddit also stores historical data which are easily retrievable from the online platform.

Thus, our objective was to use qualitative analysis methods to investigate whether social media data could have been a viable data source to understand COVID-19 symptoms early on in the pandemic and to present patients' perspective of their COVID-19 symptoms. We analyzed patient-reported COVID-19 experiences from the social media website Reddit by creating user-reported symptom profile, testing experiences, and medication use during the early phase of the COVID-19 pandemic.

2 | METHODS

COVID-19 patient-reported data were extracted, filtered, reviewed, and coded from social media to examine real-world experiences of disease. The community forum, COVID-19 positive, within Reddit was selected for this project. This community forum was created in March 2020 and is described as "A place for people who came back positive for COVID-19 can share your stories, experiences, answer questions, and vent!"

Content and metadata for original posts and comments made in the subreddit between March to August 2020 were collected using

TABLE 1 NVIVO codebook

Code	Definition
Symptoms	
Loss of senses	Any explicit mention of experiencing a loss of senses, loss of taste, or loss of smell, or descriptions indicating a loss or alteration in either taste or smell
Fever/chills	Any explicit mention of experiencing fever or chill or having a temperature measurement above 99.5°F. Note: these have been combined to reflect CDC COVID-19 symptom list
Headache	Any explicit mention of experiencing headache or migraine
Ache/pain	Any explicit mention of experiencing ache, body ache, muscle ache, or other types of pain including sharp, stabbing, burning, etc. excluding headache
Fatigue	Any explicit mention of experiencing fatigue or description of increased physical or mental tiredness
Cough	Any explicit mention of experiencing cough
Nose/congestion	Any explicit mention of experiencing runny nose or congestion, in addition to presence of other discomfort of symptoms related to the nose. This excludes chest congestion.
Throat	Any explicit mention of experiencing sore throat, or other discomfort in the throat including mild discomfort without pain
Shortness of breath (SOB)	Any explicit mention of experiencing shortness of breath, including descriptions of difficulty breathing. This excludes SOB attributed by the user to other conditions such as heartburn or anxiety.
GI	Any explicit mention of experiencing nausea, vomiting, and diarrhea, in addition to other gastrointestinal symptoms such as heartburn.
Chest/lungs	Any explicit mention of experiencing symptoms relating to the chest or lung area. This excludes objective heart measurements such as blood pressure or tachycardia.
Cog	Any explicit mention of experiencing cognitive deficits, including brain fog, confusion, or inability to concentrate.
Eye	Any explicit mention of experiencing symptoms related to the eye
Skin	Any explicit mention of experiencing symptoms related to the skin
Dizziness	Any explicit mention of experiencing dizziness or inability to balance
Off and on symptoms	Descriptions of situations in which symptoms presented, resolved, and were then experienced again by an user. This includes over any time frame including less than 1 day to multiple weeks with descriptions of relapse.
Nonspecific/ allergies	Descriptions of symptoms that do not indicate specific symptoms. This includes general mention of "symptoms," or more specific descriptions like "flu-like," "allergy," or "headcold."
Exacerbations	Descriptions of exacerbations of previously controlled concurrent medical conditions by users. This includes worsening of symptoms or increased need for medication therapy.
Testing	
Confirmation	Statements indicating an user received a COVID-19 diagnostic test and received a positive result
Repeat test	Statements indicating an user received more than one COVID-19 diagnostic test
Antibody	Statements indicating an user received a positive COVID-19 antibody test
Reason for testing	Statements indicating there was a specific motivating factor for receiving a test, both self-driven or as required for external reasons
Medication	
Antibiotics	Mentions of antibiotics prescribed to and taken by users
Inhalers	Mentions of inhalers prescribed to and taken by users, including steroid inhalers
Steroids	Mentions of steroids prescribed to and taken by users, including steroid inhalers
Acetaminophen	Mentions of use of brand or generic products containing acetaminophen
NSAIDs	Mentions of use of brand or generic products containing NSAIDs
Cough	Mentions of use of brand or generic products indicated or used for cough
Combination	Mentions of use of brand or generic products used for multiple indications
Sleep	Mentions of use of brand or generic products indicated or used for sleep
GI	Mentions of use of brand or generic products indicated for dyspepsia
Vitamins/herbals	Mentions of use of vitamin or herbal products
Other	Mentions of medications or drugs not coded under any other medication code

(Continues)

TABLE 1 (Continued)

Code	Definition
Health-related quality of life	
Impacts on daily living	Descriptions of activities or tasks users were unable to perform or unable to perform at the same capacity as prior to their infection
Post-activity relapse	Descriptions of a return or worsening of symptoms following a period of moderate to intense activity or exercise.

the Pushshift Application Programming Interface (API)[11] with R (R Core Team, 2021) version 3.6.2. R package redditr[12] was used to collect comments and posts from the subreddit by interfacing with Pushshift API. Information used from each post included the user's name, post body, and date posted. Original posts also included a post title and a self-assigned tag or "flair" category by the users, based on topic, for example, "Tested Positive," "Presumed Positive," or "Questions-to those who tested positive."

Data for two quarters, March to May 2020 (Q1) and June to August 2020 (Q2), were retrieved. To reduce data volume, we selected posts tagged by users as "Tested Positive" or "Tested Positive-Me" and/or users who posted at least 30 times within any calendar month indicating significant engagement in the community. Exclusion criteria included automated posts, deleted or removed posts, and posts from users who had a moderator role in the community, or users who explicitly stated they were outside the United States. All posts from users in the remaining dataset were reviewed to determine if a user had confirmation of diagnosis through testing. Posts were categorized as Tested Positive when the user's comments indicated at least one positive COVID-19 and/or antibody test; Tested Negative when the user's comments indicated a negative COVID-19 test without ever having a positive COVID-19 and/or antibody test; Unconfirmed/Questionable/Presumed if the user did not have a confirmed positive or negative COVID-19 or antibody test, but reported experiencing COVID-19 symptoms; and Commenter Only where COVID-19 experience or testing status was reported by a third party or users who wanted to engage in the community.

User case files (see data supplement) were created for all users using a standardized template, capturing the username, number of posts, demographic information, case progression, testing information and medication or treatments used. Original posts and comments were uploaded to the data analysis software program, NVivo for qualitative coding. Prior to data analysis, we created a codebook to code concepts reported by the users. Codes and definitions were created based on both known information about user symptoms, user testing, medication use and updated iteratively during the data analyses (Table 1). User case files were then analyzed for information about symptoms, testing experiences, medication or therapies, and quality of life impacts of disease. The study was exempt from Institutional Review Board (IRB) review because user data are completed anonymized and users cannot be identified; the data are freely available for query and the study was conducted as public health surveillance activities.

3 | RESULTS

In Q1, there were 93 982 original posts and comments from 15 625 users. In Q2, there was an increase to 114 910 original posts from 19 922 users. After applying selection and exclusion criteria, the dataset was reduced to 31 759 posts and comments from 720 users in Q1 and 40 446 posts and comments from 1649 users in Q2. Majority of users were excluded due to low activity (less than 30 posts). The final dataset reviewed included only users categorized as "Tested Positive" which included 1296 users (280 in Q1 and 1016 in Q2).

3.1 | Demographic data

Across both quarters, demographic information was not consistently reported by users, 581 and 443 users reported their age and gender respectively; therefore, we summarized data for those who reported this information (Tables A1 and A2). Over 90% were under the age of 40 years, with the majority in their 20s. There were slightly more females (54%) than males (46%) among the 443 people who reported their gender. There was representation of at least one person from 36 states, with the majority coming from states with higher populations. New York (NY), California (CA), and Florida (FL) were the more frequently reported states across both quarters. Other commonly reported states for Q1 and Q2 were New Jersey (NJ), Massachusetts (MA) and Texas (TX), Arizona (AZ), respectively.

3.2 | Chronological display and frequency of reported symptoms

In Figure 1, we present COVID-19 symptoms for selected users highlighting the reporting of these symptoms before naming by the CDC. For both quarters, we also summarized the frequency of reported symptoms (Figure 2). After the first case of person-to-person transmission on January 30, 2020, the Centers for Disease and Control (CDC), on February 23, 2020, announced that COVID-19 symptoms were similar to those of influenza (fever, cough and shortness of breath [SOB]).[13] On April 27, 2020, the CDC added six more conditions that may be associated with COVID-19; chills, repeated shaking with chills, muscle pain, headache, sore throat and new loss of taste or smell.[14] A third update was made on July 27, 2020 adding five more symptoms: nausea, vomiting, abdominal pain, diarrhea and fatigue.[8] Our review of Reddit data reveals that users were reporting more in-

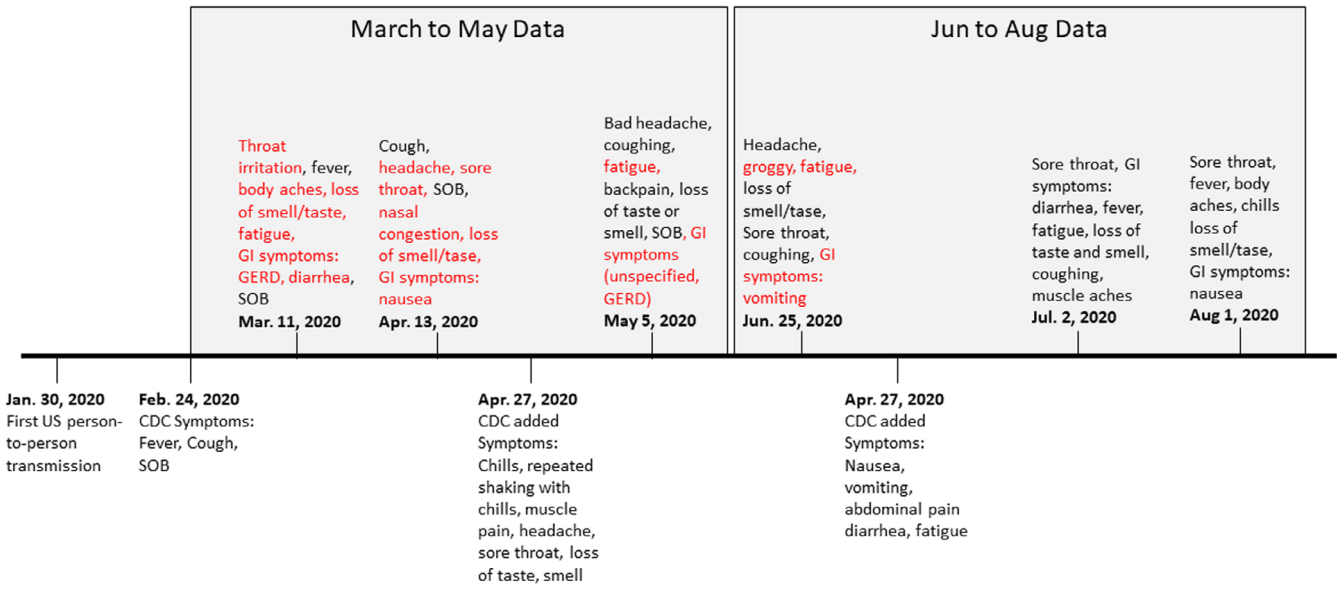

FIGURE 1 Chronological display of reported COVID-19 symptoms from selected users and sequence of the CDC's release of COVID-19 symptoms. The timeline presents COVID-19 symptoms reported by selected users (dated) with a chronological display of the Centers for Disease Control (CDC) updates to the COVID-19 symptom list. In red are COVID-19 symptoms reported by users but not named as a CDC symptom in the time frame the users reported the symptom. *SOB, shortness of breath; GI, gastrointestinal

depth description of COVID-19 symptoms in March and April 2020. As shown in Figure 1, users reported gastrointestinal (GI) symptoms: nausea, diarrhea, GERD; loss of smell or taste, fatigue, muscle pain, headaches before these symptoms were named by the CDC.

Our review of the data in Q1 revealed that in line with the known symptoms at that time, users reported fever/chills, loss of sense of smell and taste, and aches more frequently than other COVID-19 symptoms (Figure 2). We also observed that users were reporting frequently, GI symptoms (31.4%); fatigue (38.6%) and nasal symptoms (23.6%) in Q1 before these symptoms were added to the CDC symptom list. In Q2, loss of sense of taste or smell (38.8%), aches or pains (32.9%) and fever/chills (17.0%) were the top three frequently reported CDC-listed symptoms (Figure 2).

3.3 | Qualitative symptoms

Many users used basic keywords and common descriptors to discuss their symptoms, but there were some symptoms for which users provided richer descriptions. Symptoms such as fever and cough tended to be described plainly (e.g., temperature readings for fever; dry vs. wet or productive vs. non-productive cough), whereas other symptoms, such as chest symptoms and pain, were often described with detail and nuance. While some users described a short course of disease with acute symptoms followed by complete resolution, others described non-linear progression of illness and recovery with unresolved or relapsing symptoms. In the following sections, we present description of COVID-19 symptoms along with illustrative quotes from users.

3.3.1 | Loss or attenuation of the senses of taste and smell

The CDC published "Loss of Senses" as a symptom as part of its April 2020 update, and it was frequently reported as a hallmark symptom of the disease. However, as early as March, Reddit users reported not only a complete loss of smell and/or taste but also an alteration of the senses and/or an attenuation or loss of the senses. Some users explicitly described their inability to taste and smell ("it's slow coming back. When I lost it, it was like turning off a switch. Coming back, is like a dimmer"; "…. I lost my smell and taste the third week (around day 16) ….It was obliterated, completely gone, for perhaps 9 days?" "I'm not sure…It started verrry slooowly returning; NO SENSE OF SMELL AT ALL,") while others described the ability to discern tastes like salty or sweet but not the distinctive flavors of food or drink" ("my taste was limited to being able to discern salty, sweet, and spicy, but not actual flavors.;" "there are some flavors that my palette can pick up on by themselves but it can't register a complex combination of them. for example, i can taste cheese by itself but if i make like a cheese quesadilla with chicken i can't taste the concert of flavors and it's just like chewing on texture with no flavor at all."). In some cases, users experienced phantom taste and smell perception. These included reports of a "tutti frutti" smell, cigarette smell, and metallic taste in the mouth, despite having nothing in the mouth ("I smell a cigarette smoke smell constantly. It's driving me crazy because I still have no real sense of smell or taste. Just this phantom smoke smell"; "Lost my sense of smell 3 weeks ago. It started with a bleachy odor and then smell went out completely." "Taste never left completely; but got very general with sweet/sour/salty/Bitter"). Duration of this symptom ranged from a few days to multiple weeks, with many users indicating they only

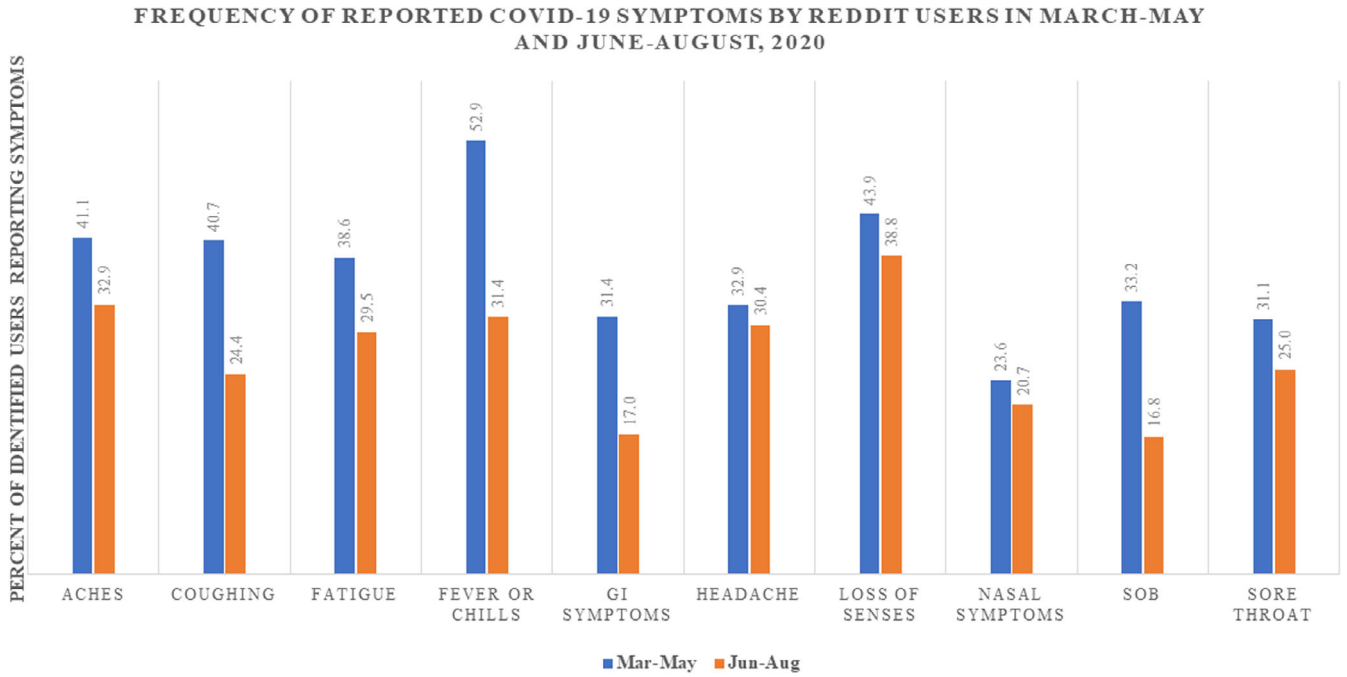

FIGURE 2 Frequency of Reported COVID-19 Symptoms by Reddit Users in March–May and June–August, 2020 Sore throat included sore throat and other throat-related symptoms (throat irritation, tickling in the throat). Nasal symptoms included nasal congestion, runny nose, blocked nose. GI symptoms included nausea, vomiting, diarrhea, abdominal pain, gastro-esophageal reflux disease (GERD)

partially or gradually recovered. There were in-depth descriptions of various smell and taste tests involving different household goods and products, including coffee, cleaning products, essential oils, and spicy foods. Changes in appetite due to changes in smell and taste were also reported.

3.3.2 | Chest and lung symptoms

The CDC COVID-19 symptom list published May 26, 2020, did not include chest or lung related symptoms but advised that "persistent pain or pressure in the chest" is an emergency warning sign for which emergency medical care should be sought. Some Reddit users reported chest pains and difficulty breathing that sometimes resulted in hospitalization, while others experienced milder chest-related symptoms such as minor discomfort, chest "congestion" or "tightness." Some users who sought medical care for these symptoms received diagnostic testing including X-rays and/or CT scans and received secondary diagnoses of pneumonia while others had clear images while symptomatic ("I felt like i had a dumbbell sitting on my chest and my heart rate would go to the 120-130 when standing and trying to do things. They did chest x ray and ekg and said it was fine"). Chest pain or tightness was not always associated with shortness of breath or coughing, with some users explicitly reported a lack of difficulty breathing ("My chest feels kinda heavy and sore-like someone punched me and my diaphragm area also feels sore. All of that said, I have no real shortness of breath"). Chest-related symptoms were also sometimes described as being related to other health-related issues. Some Reddit users reported chest-related symptoms associated with other conditions, such as heartburn or anxiety. These appeared to resolve with monitoring and symptomatic treatment. Reddit users also generally described different types of pain and irritation in the lungs ("I get the lung burning again every time I exercise or talk for more than about half an hour (work calls suck)."), including some accounts of a burning or itching sensation. Some Reddit users only described a feeling of irritation with suggestions that they felt their lungs were "inflamed" ("Lungs started to feel "irritated." It wasn't difficult to breath, but I could tell my lungs were inflamed").

3.3.3 | Ache or pain symptoms

In its April 2020 update, the CDC added "muscle pains" to the COVID-19 symptom list; this was later revised to "muscle or body aches." Although Reddit users reported aches and body aches, they also described the location of the pain, the quality of the pain "(e.g., sharp or stabbing, burning or nerve pain)" and other symptoms associated with the pain beyond just generalized aching ("awful burning in my joints…All my joints were burning horribly, knees, shoulders, hips, ankles even my toes would throw me a pain that made me exclaim out loud"; "Legs hurt REALLY bad. Felt like they were gonna fall off. Pain was 13/10."). For example, pain involving the joints like the shoulders, elbows and knees were sometimes described as inflammation of the joints and burning of the joints ("Like I have just recently developed a hot burning pain in my shoulder for the first time in my life."). These were sometimes associated with exacerbations of

prior joint injuries. Sharp pains were often described as associated with deep breaths or difficulty breathing ("I suddenly developed some sharp, charlie-horse like pain in my left shoulder that seemed to wrap around my left side and go near my chest. It only hurts upon breathing deeply, laughing, or quick/odd movement. Doesn't hurt when I cough or sneeze"). Internal pain like abdominal and lung pain were described with more visceral pain terms like burning or being "on fire Many users also described a distinct type of pain in the eye related to the muscles or movement in different directions. These were sometimes associated with concurrent headaches or migraines but tended to be described as more of an isolated concern. Other areas for pain included the neck and back ("Terrible neck pain. Couldn't even walk down the steps without wanting to cry from pain").

3.3.4 | General symptoms and disease progression

In addition to specific symptom reports presented above, users frequently provided more general information about their overall COVID-19 experience including nonspecific descriptions ("One day I'll feel 95% better, take a walk, do some errands. Another day I'll try to do one thing and it will send me back to bed"), in-progress or post-resolution symptom diaries, relapses, and severity of illness. Some Reddit users would report a simple acknowledgement that they were not asymptomatic while others provided more detail like having "head cold" or allergy symptoms. Symptom diaries similar to those potentially included in electronic health records (EHR) provided information regarding timing of individual symptoms and showed the range of experiences with clinical case progression. Some Reddit users described their COVID-19 disease progression and recovery as a complete resolution of symptoms after a short period of time, while others experienced waxing and waning of disease symptoms within acute phases of disease and over a chronic phase, with relapse following resolution. Users described symptoms resolving and then returning within hours or days and expressed frustration with experiencing new or returning symptoms. Some Reddit users reported initial resolution of symptoms, but later reported that symptoms returned as if they were re-experiencing acute infection. These recurrences of symptoms associated with acute infection were sometimes accompanied by repeat COVID-19 tests, with Reddit users reporting both positive and negative results after their first COVID-19-positive test.

Descriptions of overall disease severity ranged from asymptomatic and mild disease to severe or life-threatening requiring hospitalization. COVID-19-positive users also experienced exacerbations of other respiratory and non-respiratory conditions like asthma ("At this point, my asthma that has gotten extremely worse. I can't stop coughing, wheezing, chest tightness. I use my advair, and nebulizer and don't feel relief."), diabetes, hypertension ("My BP was controlled before I had covid, now I'm on 2 meds and it's still not fully under control."), attention deficit and hyperactivity disorder ("We both tested positive for COVID-19 back-to-back this week, and our ADHD feels like it's increased tenfold! We're still taking a (lower) dose of our Adderall, but ot feels like we aren't taking it at all."), muscle and joint injuries, and menstrual symptoms ("I have endometriosis and I had the worst cramps since my surgery during my COVID cycle.").

3.3.5 | COVID-19 testing (and diagnosis)

Access to COVID-19 testing was limited early in the pandemic. As access increased, users were able to confirm their diagnosis. These confirmations were expressed in different ways, with many using the phrase "tested positive." Other users indicated they got tested and later received a phone call or other contact confirming a diagnosis. There were also users who used a "+" symbol to indicate their positive COVID-19 test result. Some users who previously received a positive test for active infection also received antibody tests. Users who were tested repeatedly for COVID-19 sometimes showed a negative test indicating resolution after a positive test, while others had multiple positive tests over weeks or months with or without symptoms of the disease.

Users also indicated their reasons for being tested or consulting a healthcare provider. Some common reasons included a sudden loss of taste or smell ("Just got tested today because I've lost all smell and taste."), a requirement for employment (including employees of healthcare or long-term care facilities) ("I work at at nursing home. My job had the whole building tested on June 26th. Results came back July 5th and I was positive") or travel ("I arrived in Los Angeles from out of state on a full flight and got tested a few days later to be safe"), exposure to a COVID-19-positive patient, and known risk factors for severe COVID-19 such as comorbid asthma or diabetes ("my insulin resistance was increasing (Type 1 Diabetic) which is a sign of infection, so I got tested on the 11th").

3.3.6 | Reported treatments for symptoms of COVID-19

Users provided information about prescription and over the counter (OTC) medications, supplements such as vitamins and herbal products, medical oxygen, or other drugs for treatment of symptoms. The most frequently reported class of products across Q1 and Q2 were pain relievers/fever reducers, mostly OTC products containing acetaminophen or non-steroidal anti-inflammatory drugs (NSAID). Users who were prescribed medications were most frequently prescribed antibiotics, steroids, and/or inhalers. Antibiotics were prescribed for indications such as ear or sinus infection, confirmed pneumonia, or for prevention of secondary infection. Some users were already prescribed inhalers for conditions such as asthma, and they reported increased use of their inhalers; other users were newly prescribed inhalers for short-term symptomatic relief. Other less commonly reported prescription and OTC medications were those used to treat cough, gastrointestinal symptoms, and sleep disturbances.

3.3.7 | Health-related quality of life

Users reported both short term and long-term impacts on quality of life and ability to perform activities of daily living (ADL). Those experiencing short-term infections and those experiencing longer-term symptoms, sometimes referring to themselves as "long-haulers," reported either an inability or a reduced ability to perform tasks both inside and outside of the home. Users described increased effort needed to perform tasks such as cleaning, getting in and out of the bathtub, arising from a seated position, or carrying groceries ("Since I've been cleared, things like moving groceries, sometimes getting out of chairs, transferring to and from the car, getting in and out of the bathtub, lifting mildly heavy objects and even bending down/righting myself to load the washer/dryer now feel like they're twice as hard than they usually were."). Increased fatigue was reported following more physically demanding tasks such as lifting objects or doing yard work. Users also reported the inability to go to work or to complete schoolwork ("due to the tiredness, i can't do my finals (2 articles and several projects alongside tests for other subjects) quite as well as i wanted to"), and that effort would sometimes result in a worsening or relapse of symptoms immediately thereafter. Users reported situations in which they felt their symptoms were resolved but then relapsed following attempts to exercise. Some users expressed hesitation at attempting moderate or intense exercise based on reading about the experiences of others.

4 | DISCUSSION

Social media platforms have become avenues for users to share disease-related experiences and exchange advice on the management of their symptoms including doctor-patient interactions.[15] Due to the anonymous nature of these platforms, users often feel comfortable with providing in-depth description of their symptoms, sometimes, prior to seeking healthcare intervention to understand the symptoms they are experiencing, the outcome of the disease and what potential treatment may be effective. Such strong desire for information is likely to result in real-time reporting of symptoms which can be tapped early in an emerging pandemic to understand the disease as it evolves. Real-time or near-real-time reporting can be insightful in describing the disease symptomatology and clinical course, especially when there is an urgency to provide as much information as possible to the public.

In this study, users gave frequent accounts of their COVID-19 experiences ranging from general acknowledgement of symptoms to in-depth descriptions of symptom quality, location, duration, and impact. Our review of user experiences, identified COVID-19 symptoms, reported by users prior to their inclusion in the list of symptoms published by the CDC. Loss of sense of smell or taste were reported by users in early March, prior to the inclusion of this symptom in April and GI related symptoms and fatigue were reported in the March to May data, before they were added as a COVID-19 associated symptom in July 2020.

Advanced analytical approaches such as natural language processing and machine learning techniques[16] have been used to process various social media platforms.[17–19] Although, these advanced techniques can effectively extract "themes" and "terms" from a large volume of data, the capture of in-depth descriptions of patient-reported symptoms is often lost with automated processing of the data. Further refinement of these analytical approaches to reflect the data is possible but may be difficult to implement. In our study, we have attempted to preserve reported symptoms as described by the user to present a clear clinical picture of the symptoms from the patient's perspective. This type of information is desirable when there is limited knowledge of a disease early in a pandemic. In some instances, tagging "themes" or "terms" from the data may not be sufficient to characterize disease. For example, while the CDC described the COVID-19 symptoms as a loss of smell and taste, users frequently described an attenuation rather than complete loss of the senses. In some cases, phantom smells and tastes, rather than losses, were reported. "Trouble breathing" and "persistent pain or pressure in the chest" were listed as emergency warning signs for which to seek healthcare, but users did not report seeking medical care, which could indicate milder symptoms, or unawareness of the seriousness of their symptoms. Early in the pandemic, these descriptions would have been useful to aid diagnosis or identify patients for testing when test kits were limited.

Social media has also been found to penetrate the population regardless of education, race/ethnicity, or health care access.[20] For early surveillance in a pandemic and in the context of COVID-19, social media data would have been particularly useful, especially since early data suggested racial and socioeconomic disparity in the provision of healthcare. Social media data may also be useful in understanding symptoms among younger healthier patients who are less likely to be hospitalized and may also be a source of real-time reactions to public health policies or actions such as increases in access to testing and/or immunization.

Although, the use of prescription medications may be better captured by formal healthcare data sources like EHRs and medical claims data; over the counter medications, vitamin and herbal supplements, or other non-prescription therapies may be better captured through other data sources such as social media. The use of pain relievers and fever reducers were widely reported along with cough medications or combination products and supplements like Vitamin C and D. These data might also be useful predictors of future drug shortages. Additional information relating to treatment reported includes regimen changes for medications or changes in therapies used to treat chronic conditions might indicate exacerbation of a previously controlled condition.

Analysis of social media data may also generate research questions and hypotheses for elucidation through additional qualitative studies, such as by conducting focus groups or interviews in patients and/or caregivers. This approach may be particularly useful for assessing health-related quality of life in various user subgroups. We found that changes in ADLs were not experienced by some users, whereas other users reported lingering and debilitating symptoms that affected

their ADLs, including effects on ability to conduct household chores, walk short distances, talk for long periods, and/or participate in school or work-related activities.

In this study, we identified several advantages to accessing self-reported patient experiences in the early phases of the COVID-19 pandemic, which may generally apply to any public health emergency. Early insights can be particularly useful for novel and rapidly evolving diseases to provide perspectives regarding clinical symptoms and course, disease outcomes, and to generate important questions for further assessment. Our study is however, not without limitations. Social media data include extensive use of acronyms and abbreviations, lack of standardization between individual data points (e.g., between posts from different users) and across data sets (e.g., from different social media platforms). We are also unable to verify individual user details and content. A prospective survey of COVID-19 patients will address this limitation, however, the challenge of recruitment especially for patients with milder forms of the disease could render this approach infeasible. The large volume of data also poses a challenge to filtering and identifying information that is relevant to the dataset. However, we have attempted to develop a framework to select active users that are likely to provide in-depth description of clinical symptoms. The conversational nature of social media can be difficult to follow from a data analysis perspective if the original posts are not in proximity to extended posts. Lastly, reddit users in the study are likely younger with less severe disease and may not be generalizable to all COVID-19 patients.

5 | CONCLUSIONS

Social media can potentially serve as an early surveillance data source in a pandemic and provide early insights into several aspects of disease experience and medication use. For the COVID-19 pandemic, early clinical information in the COVID-19 pandemic on presentation of severe disease cases or from a convenience sample limited the generalizability of these data. Patients with less severe cases who interacted on social media might have offered information about earlier signals of disease. In-depth discussions of symptoms can also aid in broader understanding of the natural history of the disease and a clinical presentation of symptoms from the patient's perspective.

6 | RESEARCH IN CONTEXT

6.1 | Evidence before this study

Several studies have reviewed social media to identify predictors of COVID-19 cases, describe government interactions and responses to the public, examined mental health and assess infodemic during the pandemic. However, limited data are available on the use of social media as a potential early surveillance data source for understanding patient symptomatology following COVID-19 infection. In this study, we extracted Reddit social media posts for users tagged as positive to explicitly describe their COVID-19 symptoms, testing, medications, or treatments.

6.2 | Added value of this study

We have created and reviewed individual case profiles of COVID-19 patients who were tagged as test positive in the Reddit social media forum. In-depth description of symptoms provides unique patient perspectives on the different types of known and unknown COVID-19 symptoms. In some instances, the "named" symptom may present in a variety of ways, of different degrees of severity and duration. Some symptoms appeared to be reported before communication to the public.

6.3 | Implications of all the available evidence

Social media data can be a resource for understanding early symptoms of novel diseases and may serve to supplement data from interactions with the healthcare systems. In situations, where little is known about an emerging epidemic, social media can guide generation of early hypotheses about the disease. Given the anonymous nature of platform, patients may more freely express themselves, to allow researchers early access to real-world information about the disease.

ACKNOWLEDGMENT

This project was supported in part by an appointment to the Research Participation Program at the U.S. Food and Drug Administration administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data used in the study are available in the supplement.

ORCID

Wei Liu https://orcid.org/0000-0002-0434-4674
Efe Eworuke https://orcid.org/0000-0001-8971-9748

REFERENCES

1. Salathé M, Bengtsson L, Bodnar TJ, et al. Digital epidemiology. *PLOS Comput Biol*. 2012;8(7):e1002616. doi:10.1371/journal.pcbi.1002616
2. Tsao S-F, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. *Lancet Digital Health*. 2021;3(3):e175-e194. doi:10.1016/S2589-7500(20)30315-0
3. Chen J, Wang Y. Social media use for health purposes: systematic review. *J Med Internet Res*. 2021;23(5):e17917. doi:10.2196/17917
4. Jiang T, Osadchiy V, Mills JN, Eleswarapu SV. Is it all in my head? Self-reported psychogenic erectile dysfunction and depression are

common among young men seeking advice on social media. *Urology*. 2020;142:133-140. doi:10.1016/j.urology.2020.04.100
5. Guidry J, Zhang Y, Jin Y, Parrish C. Portrayals of depression on Pinterest and why public relations practitioners should care. *Pub Relat Rev*. 2016;42(1):232-236. doi:10.1016/j.pubrev.2015.09.002
6. Koutrolou-Sotiropoulou P, Lima FV, Stergiopoulos K. Quality of life in survivors of peripartum cardiomyopathy. *Am J Cardiol*. 2016;118(2):258-263. doi:10.1016/j.amjcard.2016.04.040
7. Killerby ME, Link-Gelles R, Haight SC, et al. Characteristics associated with hospitalization among patients with COVID-19—metropolitan Atlanta, Georgia, March–April 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:790-794. doi:10.15585/mmwr.mm6925e1
8. Kim L, Whitaker M, O'Halloran A, et al. Hospitalization rates and characteristics of children aged <18 years hospitalized with laboratory-confirmed COVID-19—COVID-NET, 14 states, March 1–July 25, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:1081-1088. doi:10.15585/mmwr.mm6932e3
9. Ellington S, Strid P, Tong VT, et al. Characteristics of women of reproductive age with laboratory-confirmed SARS-CoV-2 infection by pregnancy status—United States, January 22–June 7, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:769-775. doi:10.15585/mmwr.mm6925a1external
10. Burke RM, Killerby ME, Newton S, et al. Symptom profiles of a convenience sample of patients with COVID-19—United States, January–April 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:904-908. doi:10.15585/mmwr.mm6928a2
11. Baumgartner J, Zannettou S Keegan B, Squire M, Blackburn J. The Pushshift Reddit Dataset. ICWSM. https://ojs.aaai.org/index.php/ICWSM/article/view/7347. Accessed January 18, 2022.
12. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2022. https://www.R-project.org/. Accessed January 18, 2022.
13. Jernigan DB. Update: public health response to the coronavirus disease 2019 outbreak—United States, February 24, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:216-219. doi:10.15585/mmwr.mm6908e1
14. The coronavirus crisis: CDC adds 6 symptoms to its COVID-10 list. NPR; April 27, 2020; 2020. https://www.npr.org/sections/coronavirus-live-updates/2020/04/27/845321155/cdc-adds-6-symptoms-to-its-covid-19-list. Accessed January 18, 2022.
15. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*. 2013;15(4):e85. doi:10.2196/jmir.1933
16. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? *BMC Med*. 2019;17(1):11. doi:10.1186/s12916-018-1247-8
17. Ding Q, Massey D, Huang C, et al. Tracking self-reported symptoms and medical conditions on social media during the COVID-19 pandemic: Infodemiological study. *JMIR Public Health Surveill*. 2021;7(9):e29413. doi:10.2196/29413
18. Luo X, Gandhi P, Storey S, Huang K. A deep language model for symptom extraction from clinical text and its application to extract COVID-19 symptoms from social media. *IEEE J Biomed Health Inform*. 2022;26(4):1737-1748. doi:10.1109/JBHI.2021.3123192
19. Sarker A, Ge Y. Mining long-COVID symptoms from reddit: characterizing post-COVID syndrome from patient reports. *JAMIA Open*. 2021;4(3):ooab075. doi:10.1093/jamiaopen/ooab075
20. Chou W-YS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. *J Med Internet Res*. 2009;11(4):e48. doi:10.2196/jmir.1249

How to cite this article: Khashei M, Janiczak S, St. Clair C, et al. Social media for early characterization of pandemic symptoms: A qualitative analysis of patient-reported COVID-19 experiences. *Pharmacoepidemiol Drug Saf*. 2022; 1-11. doi:10.1002/pds.5564

APPENDIX A

TABLE A1 Demographic information: Age

	Count of users			
Age group	**Jun–Aug ($n = 426$)**	**Mar–May ($n = 155$)**	**Total ($n = 581$)**	**Percent total ($n = 581$)**
Teens	35	4	39	6.71%
20s	255	85	340	58.52%
30s	105	48	153	26.33%
40s	22	13	35	6.02%
50+	9	5	14	2.41%

Note: Only 581 out of the 1296 users reported age.

TABLE A2 Demographic information: Sex

	Count of users			
Sex	**Jun–Aug ($n = 316$)**	**Mar–May ($n = 127$)**	**Total ($n = 443$)**	**Percent total ($n = 443$)**
Male	150	53	203	45.82%
Female	166	74	240	54.18%

Note: Only 443 out of the 1296 users reported their gender.